Multidisciplinary Development of Platforms for Protein Identification, Expression and Control at the Single Cell Level in the Post-Genomics Era

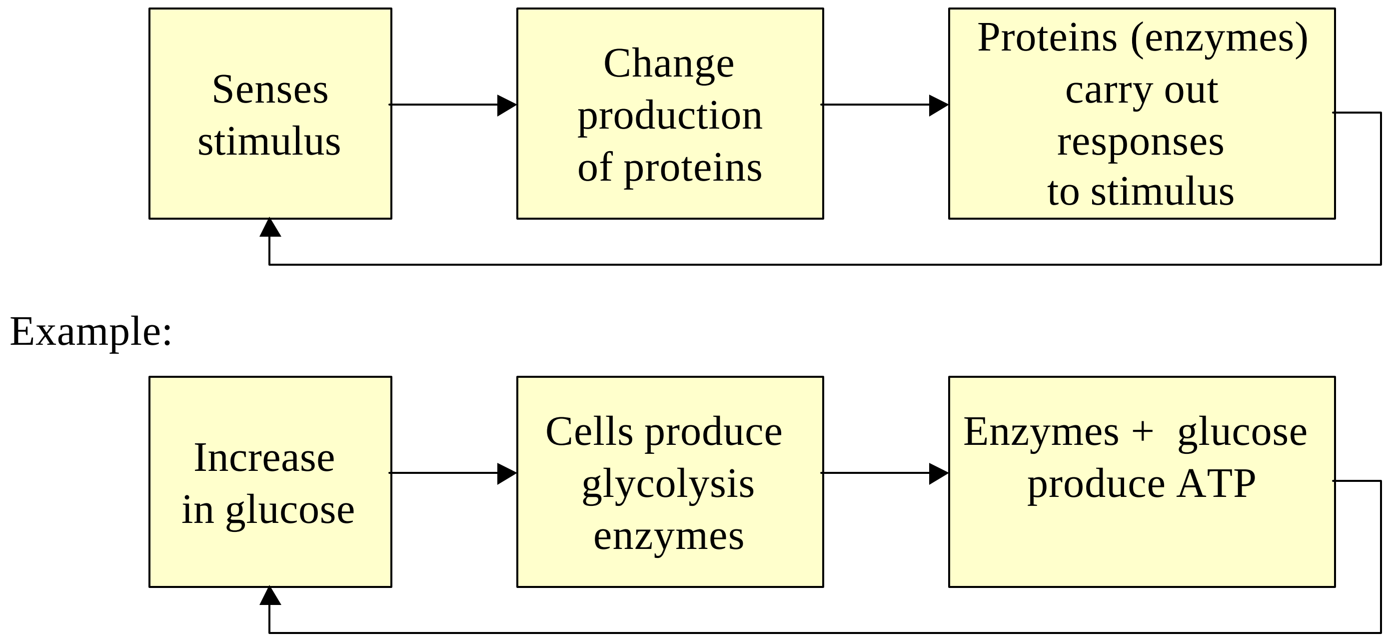
Center for Photonics and Optoelectronic Materials (POEM)
Institute for Integrative Genomics (IIG)
Princeton University

Investigators and Fields

- Nanotechnology/Elec. Eng'g:
  - Stephen Chou: 609-258-4416, chou@princeton.edu
  - James Sturm (PI): 609-258-5610, sturm@princeton.edu
- Biophysics:
  - Robert Austin: 609-258-4353, rha@suiling.princeton.edu
- Molecular Biology:
  - Edward Cox: 609-258-3856, ecox@princeton.edu
  - Jeff Stock: 609-258-6111, jstock@princeton.edu
  - Shirley Tilghman: 609-258-2900, stilghman@molbio.princeton.edu
- Bioinformatics/Computer Science:
  - Mona Singh: 609-258-2087, msingh@princeton.edu

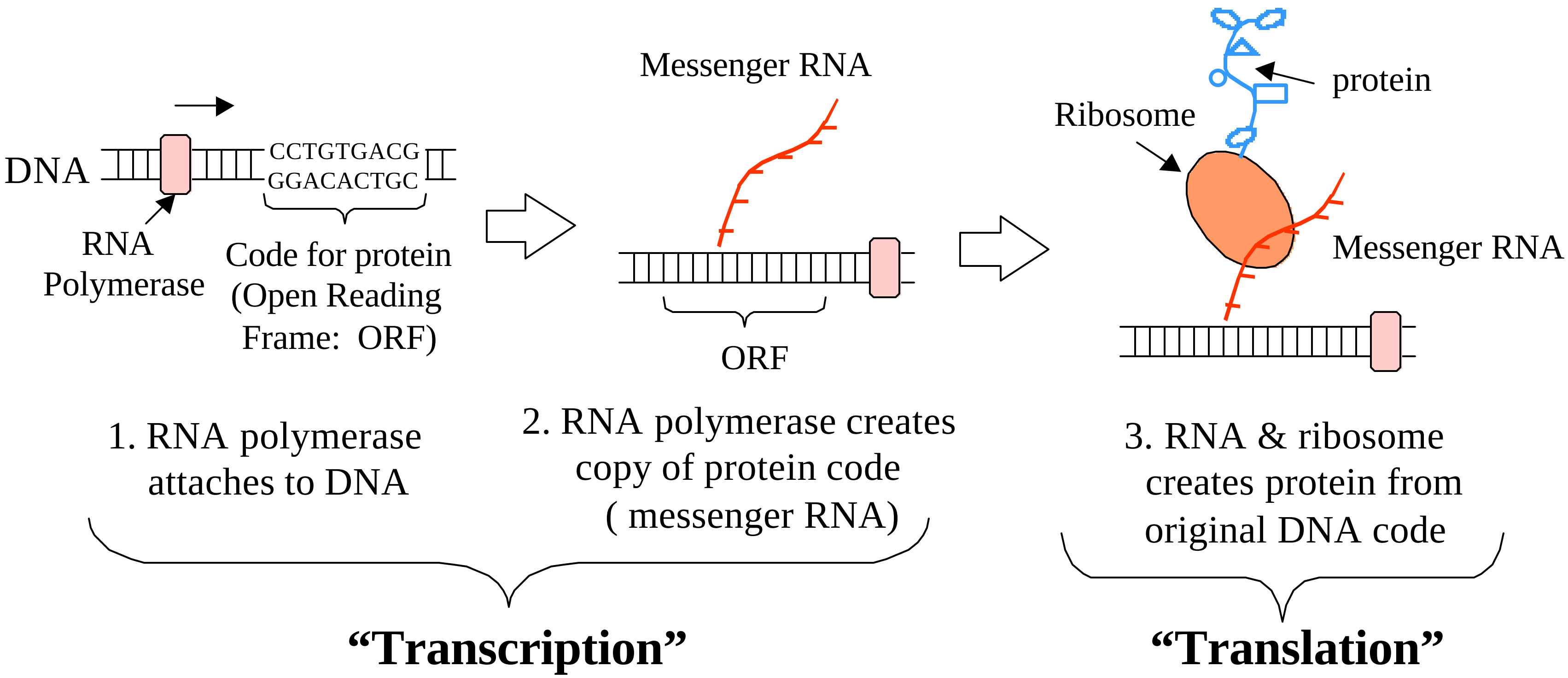
- Administrator: Kim Hegelbach, 609-258-1832, kim@ee.princeton.edu
- Industrial Liason: Joe Montemarano, 609-258-2267, jmonte@ee.princeton.edu

Regulation of Biological Organisms



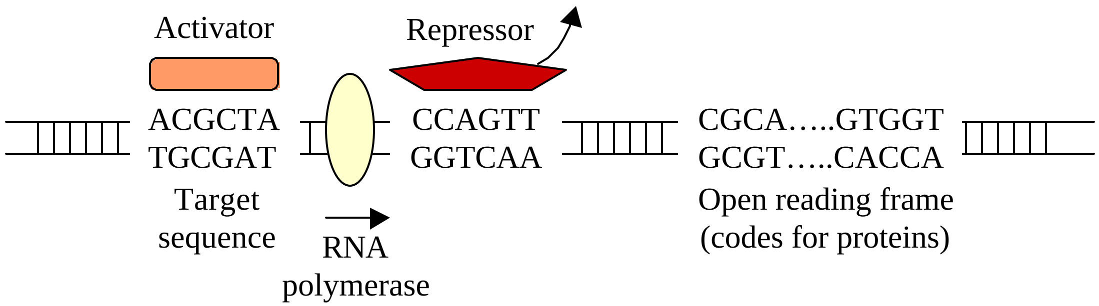
- All proteins are coded in DNA
- Central question: what controls the production of proteins?
- What is the logic which controls which proteins expressed when?

Expression of Proteins



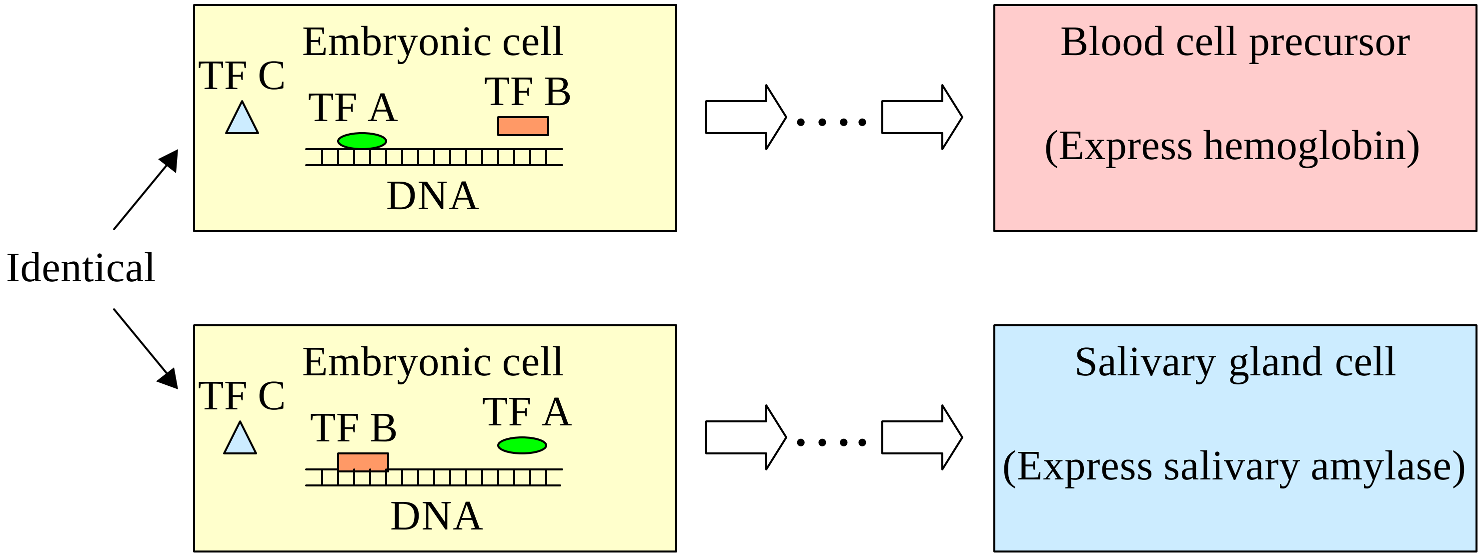
- What governs transcription?

Transcription Factors



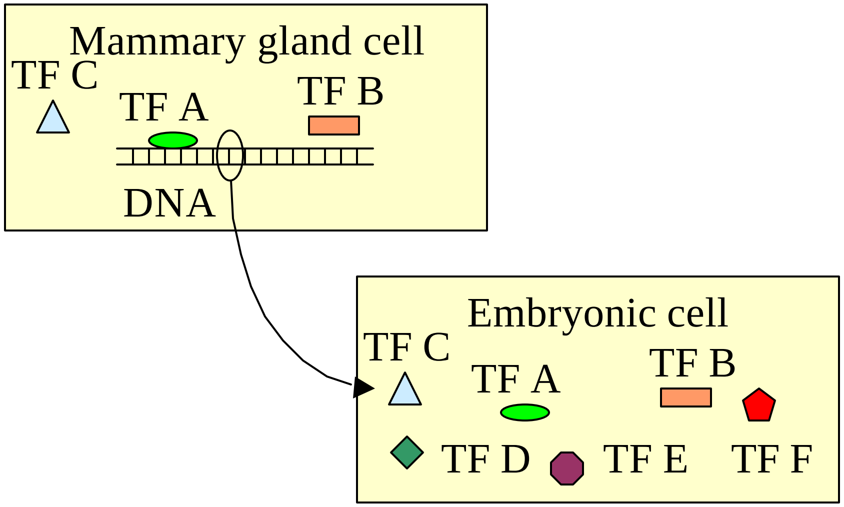
- Proteins called “transcription factors” control the transcription process
  - Affect binding of RNA polymerase to DNA
  - Affect local melting of DNA double helix
  - Affect procession of RNA polymerase along DNA
- Activators or repressors
- **Goal of this project: discover “logic” of transcription**
  - **What combination of TF’s for each protein expressed?**

Transcription Also Controls Cell Identity



- Transcription controls function of cell (creates hemoglobin vs. salivary enzyme)
- Transcription controls development of cell from embryonic cells
- Why did blood cell vs. salivary gland cell result?

Transcription and Cloning



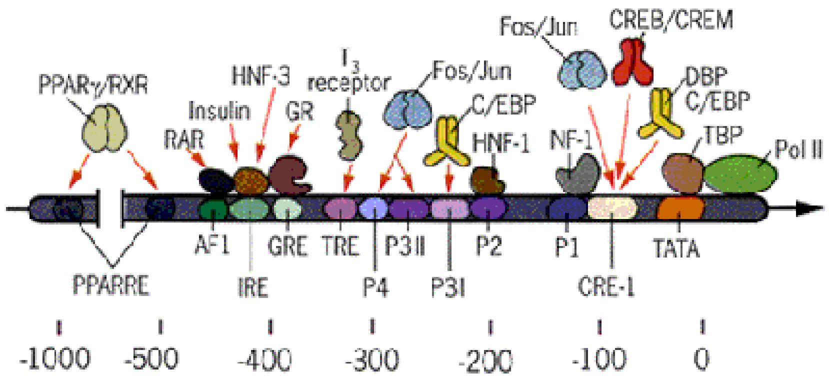
- DNA from fully developed cell plus transcription factors from embryonic cell => clone full organism
- Message is that transcription is a critical process, which can always alter fate of cell

From Basic Principles to Complex Systems

Fundamental law	*Devices/ component*	*System*
Electrodynamics	Radio	Wireless Network
Quantum Mechanics	Electrons & holes, Transistor	Computers
Machine Instructions	Programming language instructions	Software System
Gene Expression	Proteins & Biological Regulation	Engineered Biological organisms

- Genome known for many organisms including humans
- Equivalent to schematic diagram or core dump in computers
- Do not know laws which govern gene expression

Transcription of PEPCK Rat Gene



- Transcription is a very complicated process
- Heroic work of indirect experiments to unravel regulation of this *one* gene (e.g. S. Tilghman)
- Our experiment:
  - Directly examine transcription factor bound to DNA with near-field nano-scanners
  - Use bioinformatics to discern combinatorial code governing transcription

Calculate vs. Experiment?

The Theory of Everything

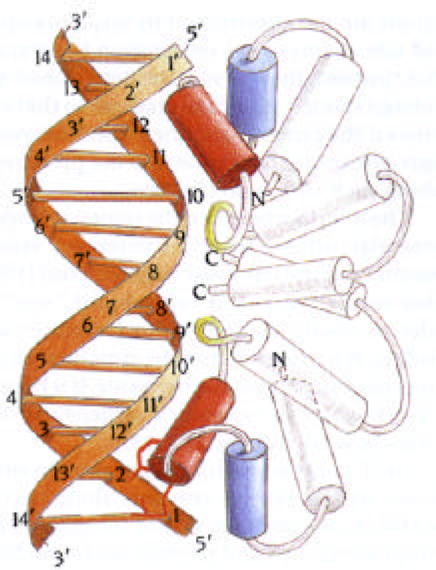
R. B. Laughlin* and David Pines[†‡§]

*Department of Physics, Stanford University, Stanford, CA 94305; [†]Institute for Oakland, CA 94607; [‡]Science and Technology Center for Superconductivity, Un Division, Los Alamos National Laboratory, Los Alamos, NM 87545

Contributed by David Pines, November 18, 1999

We discuss recent developments in our understanding of matter, broadly construed, and their implications for contemporary research in fundamental physics.

Proc. Nat. Acad. Sci, Jan., 2000

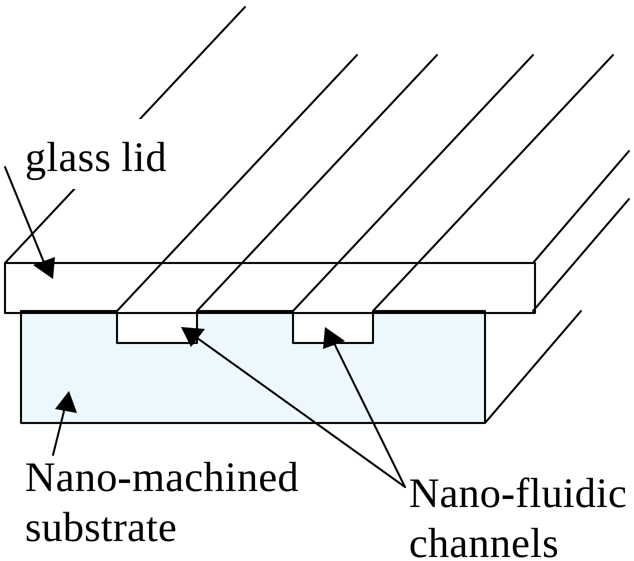


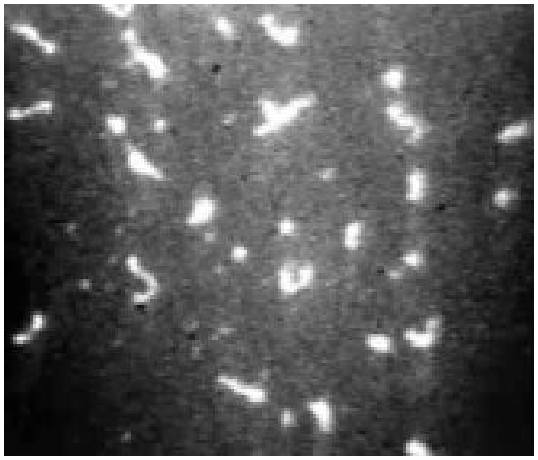
Schematic diagram of the common bacterial binding protein, the *lac* repressor, bound to DNA.

to be the least reliable precisely when reliability is most needed, i.e., when experimental information is scarce, the physical behavior has no precedent, and the key questions have not yet been identified. There are many notorious failures of alleged *ab initio* computation methods, including the phase diagram of liquid ^{3}He and the entire phenomenonology of high-temperature superconductors (8–10). Predicting protein functionality or the behavior of the human brain from these equations is patently absurd. So the triumph of the reductionism of the Greeks is a pyrrhic victory: We have succeeded in reducing all of ordinary physical behavior to a simple, correct Theory of Everything only to discover that it has revealed exactly nothing about many things of great importance.

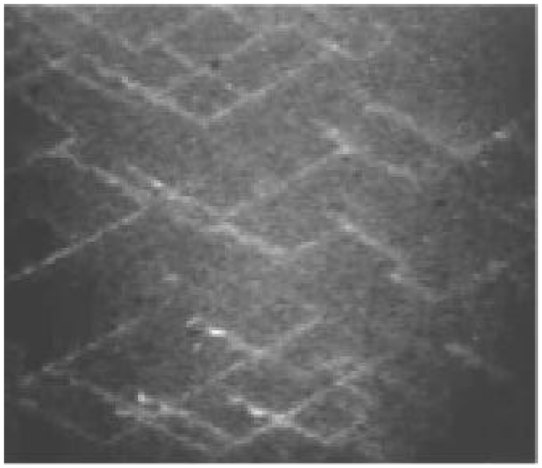
world. Horgan's book might more properly have been called the End of Reductionism, for it is actually a call to those of us concerned with the health of physical science to face the truth that in most respects the reductionist ideal has reached its limits as a guiding principle. Rather than a Theory of Everything we appear to face a hierarchy of Theories of Things, each emerging from its parent and evolving into its children as the energy scale is lowered. The end of reductionism is, however, not the end of science, or even the end of theoretical physics. How do proteins work their wonders? Why do magnetic insulators superconduct? Why is ^{3}He a superfluid? Why is the electron mass in some metals stupendously large? Why do turbulent fluids display patterns? Why does black hole formation so resemble a quantum phase transition? Why do galaxies emit such enormous jets? The list is endless, and it does not include the most important questions of all, namely those raised by discoveries yet to come. The central task of theoretical physics in our time is no longer to write down the ultimate equations but rather to catalogue and understand emergent behavior in its many guises, including potentially life itself. We call this physics of the next century the study of complex adaptive matter. For better or worse we are now witnessing a transition from the science of the past, so intimately linked to reductionism, to the study of complex adaptive matter, firmly based

Nanofluidics at Princeton (Austin and Cox)





Coiled T4 DNA (167 kbp)

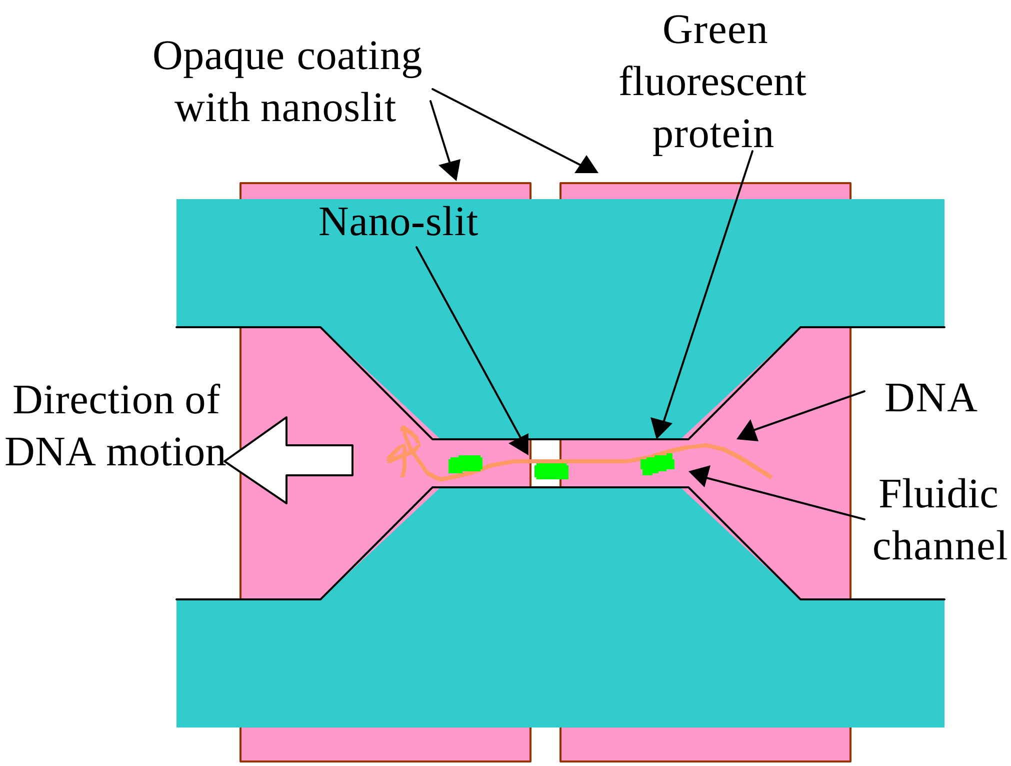


Uncoiled T4 in pulsed field

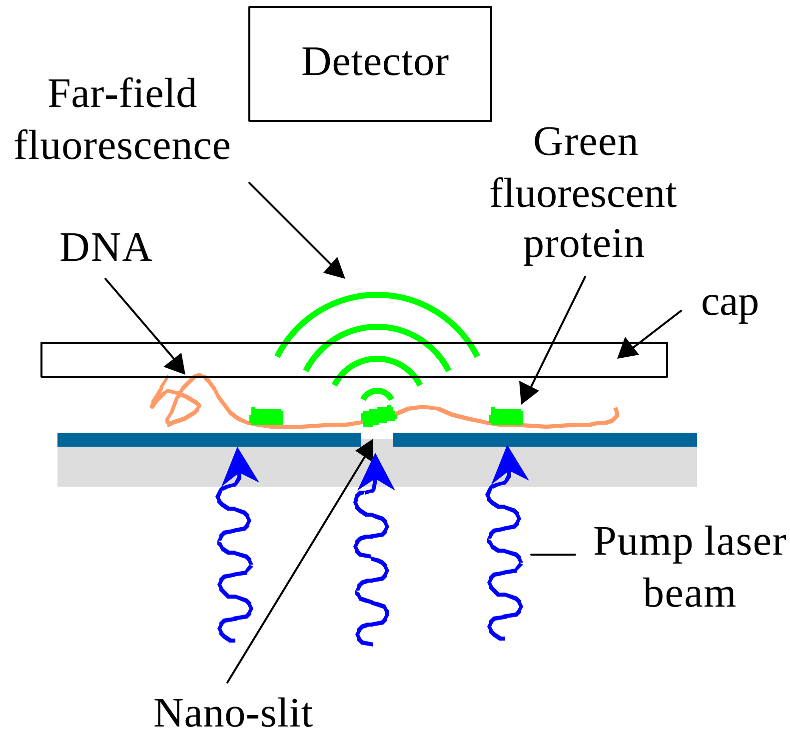
Reference: C. F. Chou et al., Electrophoresis 2000, 21, 81-90

- Lots of experience with fractionation and separation of macromolecules and cells
- DNA can be uncoiled in confined structures

Near-Field Nano Scanner

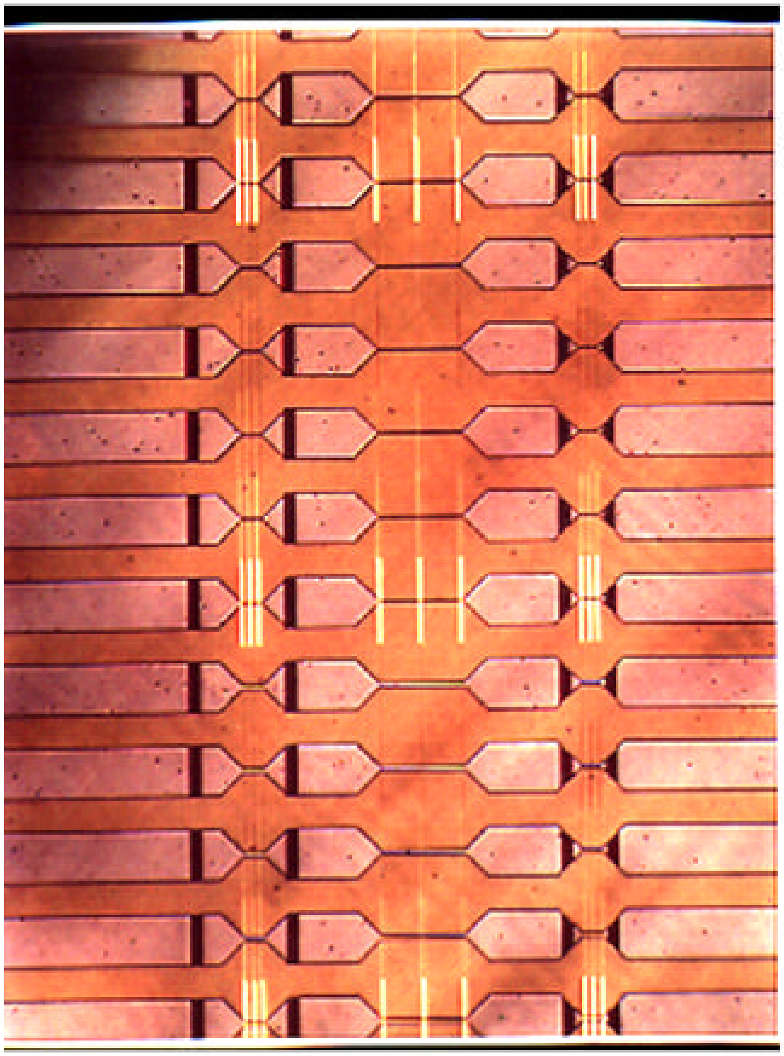


Top view



Side view

- Only fluorescent proteins above nano slit “light up”
- Only “near field” penetration of pump beam through slit
- Nanofabrication allows ~50 nm spatial resolution with optics



Optical micrograph of fabricated array

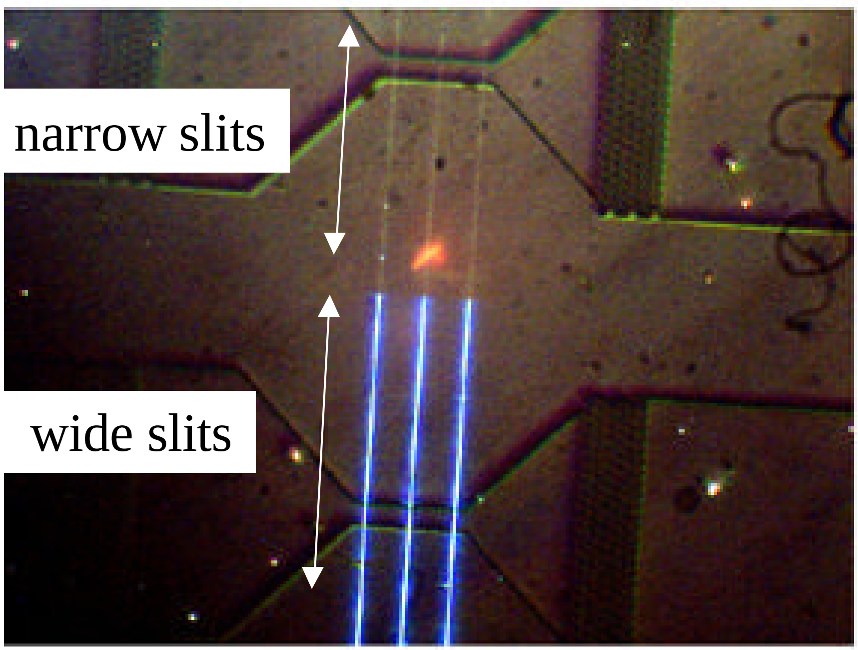
TOP illumination for structures

REAR illumination for slits:

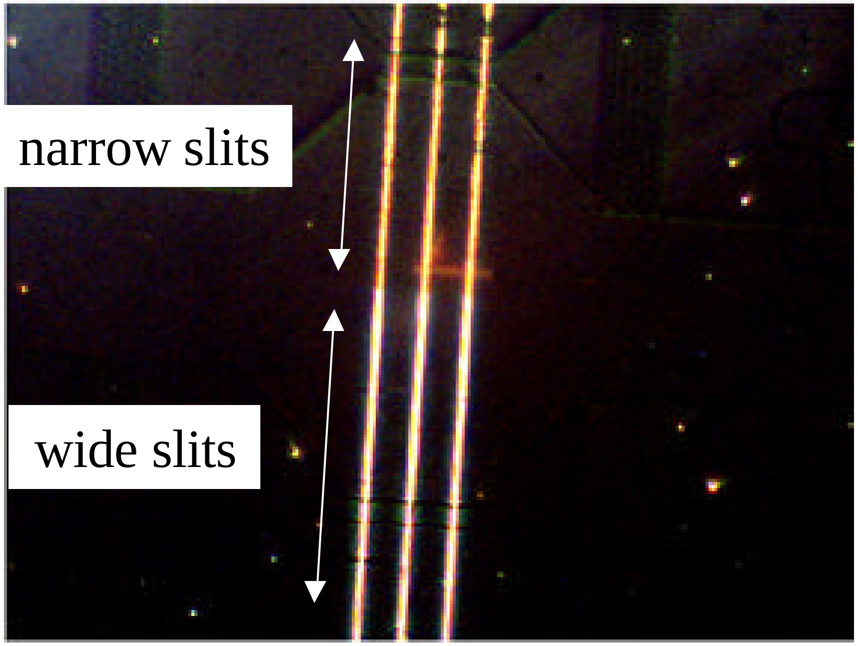
.05
.1
.3
1
2

variable width and spacing

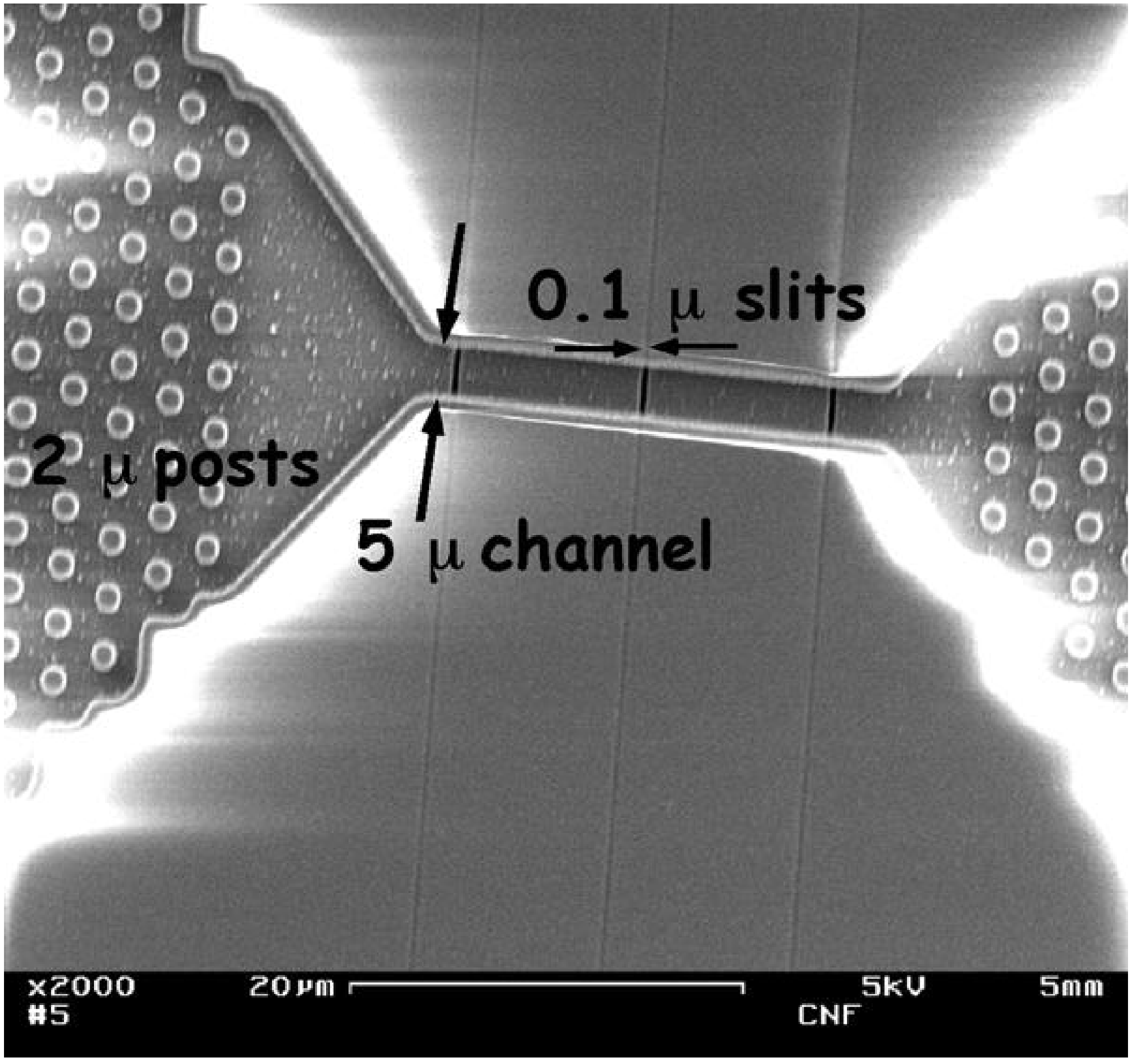
Just for fun: switching between near and far field by rotating the E polarization:



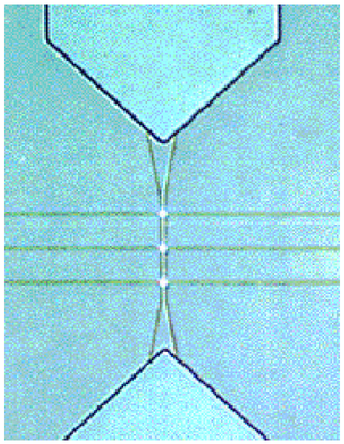
white light parallel



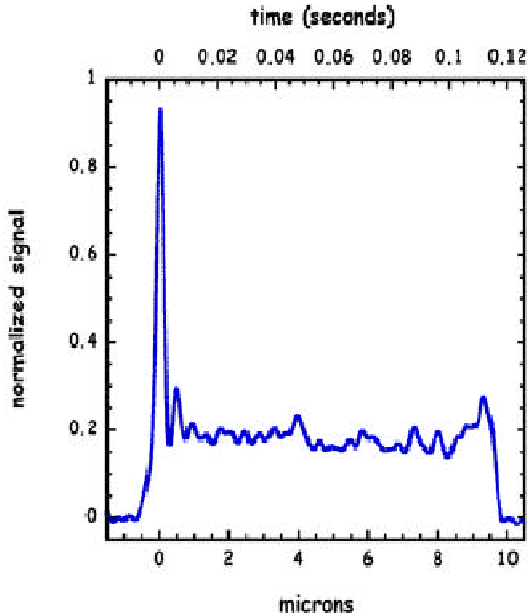
white light perpendicular


0.1 μ slits
2 μ posts
5 μ channel
x2000 20μm 5kV 5mm
#5 CNF

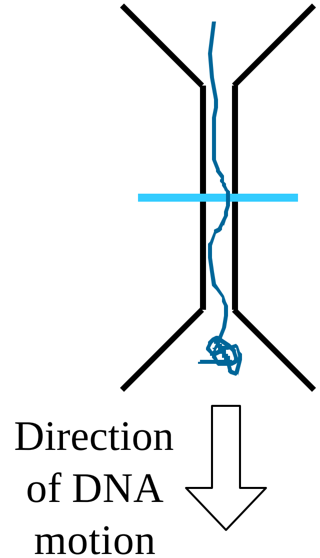
Initial Experiments



Nanofluidic channel with 100 nm slits



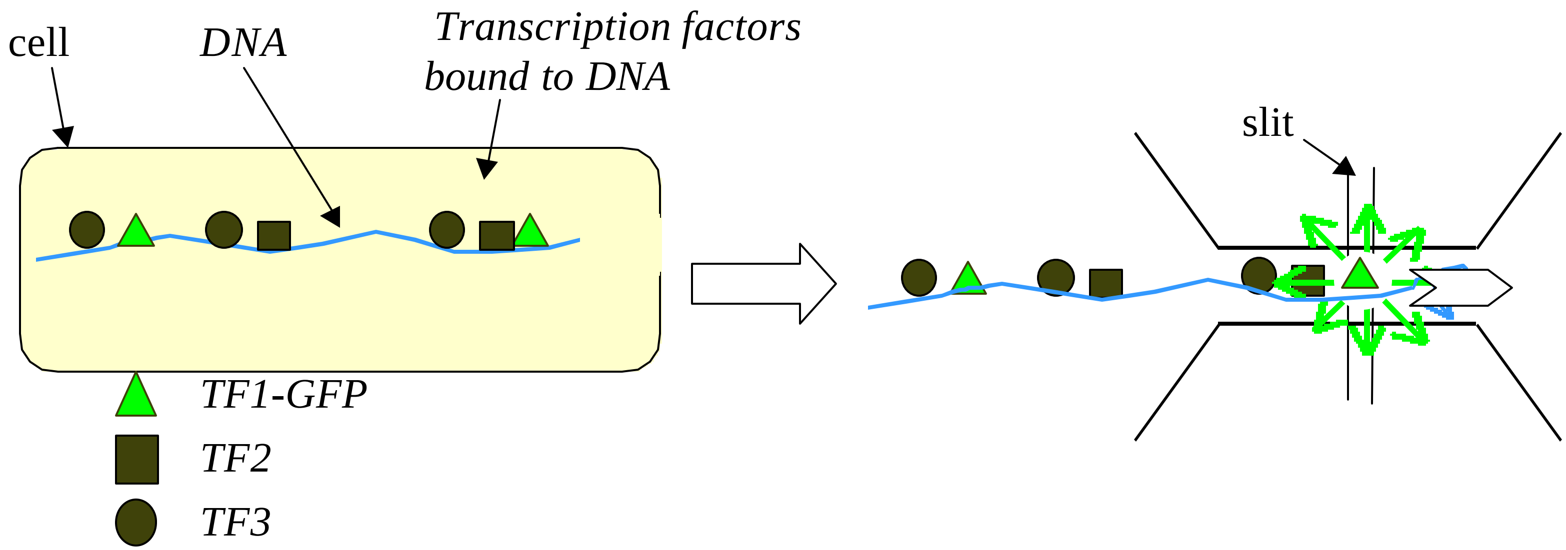
Signal vs. time as fluorescently stained DNA passes over a near-field slit



Tegenfeldt, Austin, Chan, Cox, Biophys. J. **78** (2000).

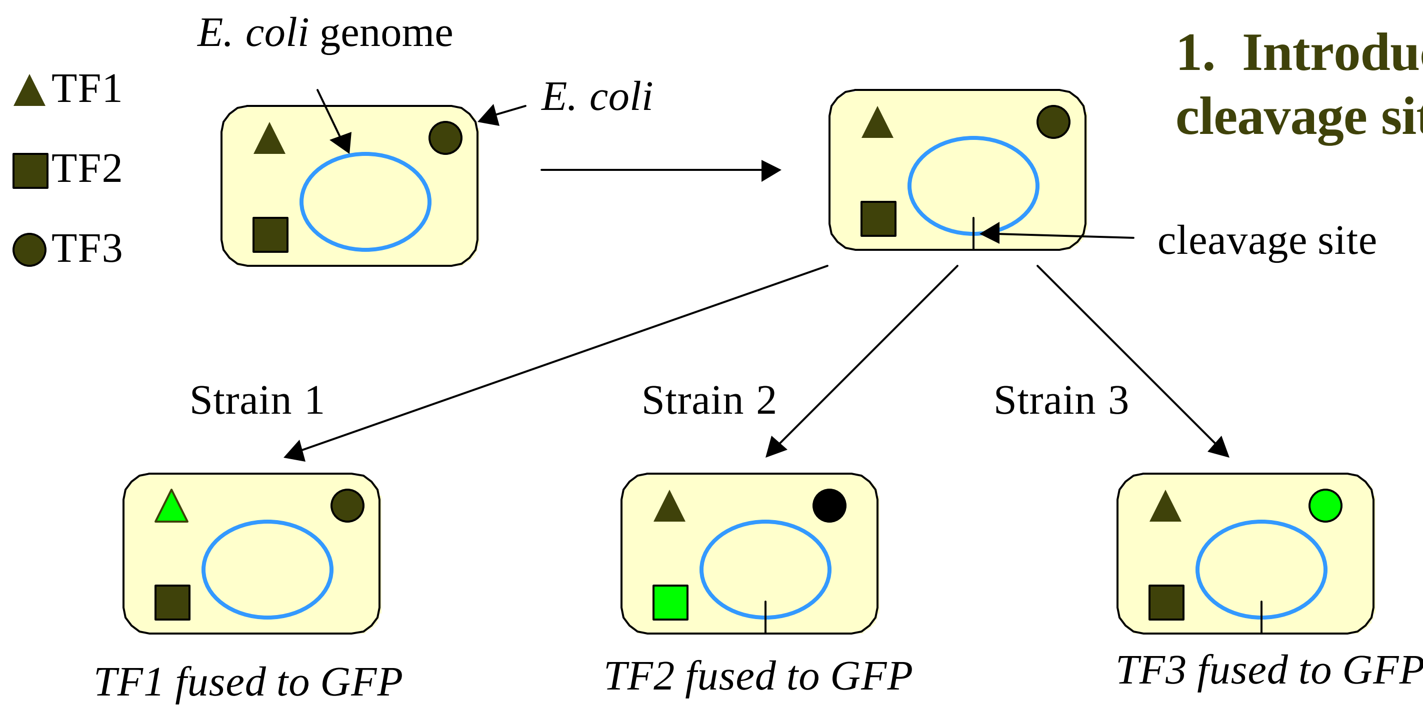
- DNA of T4 virus, fluorescent stain on entire molecule
- Large signal at leading edge: head is coiled
- Small blips from local coiling

Green Fluorescent Protein (GFP)



- Selectively fuse green fluorescent protein (GFP) to a single transcription factor of *E. coli*
- Allows one to uniquely find binding location of that TF
  - excite fluorescence of that TF only in near-field nano-scanner

Technical Program

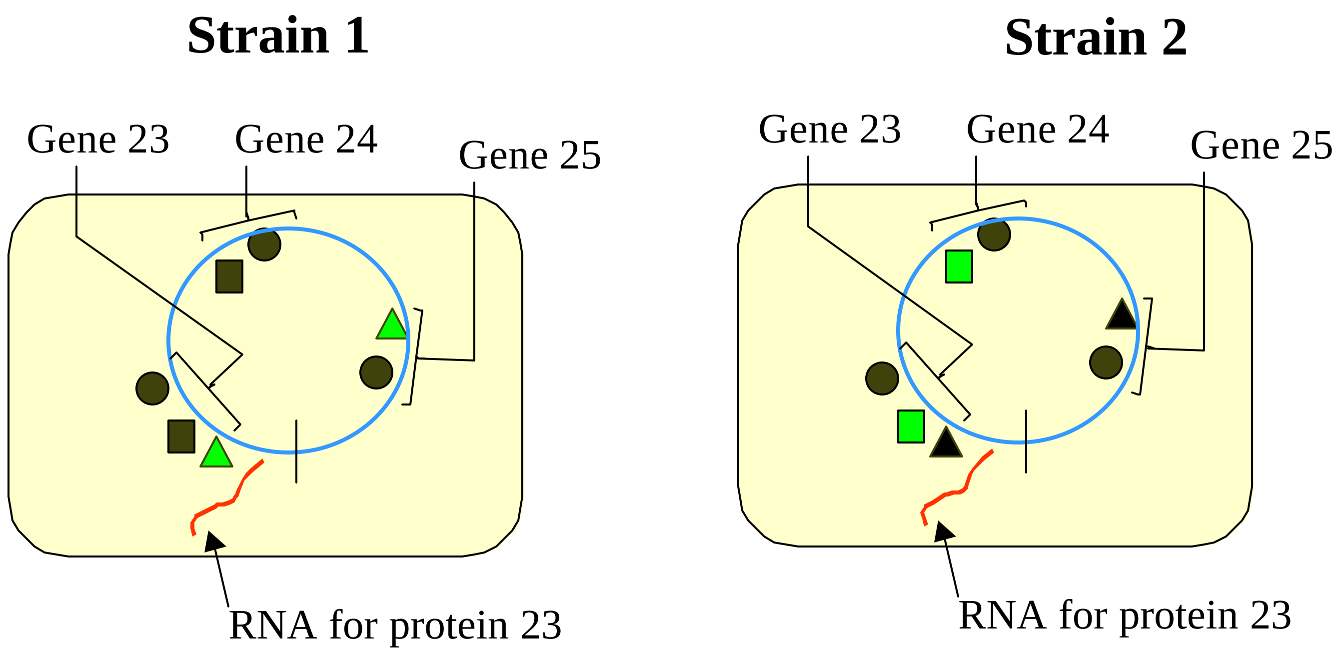


1. Introduce unique cleavage site in *E. coli*

2. Create strains, each with different transcription factor fused to GFP

- Choose *E. coli* since genome entirely sequenced, tractable problem

3. Culture all strains under identical environmental conditions



- **Detect RNA produced *(which genes are expressed?)***
  (e.g. Gene 23, but not 24, 25)

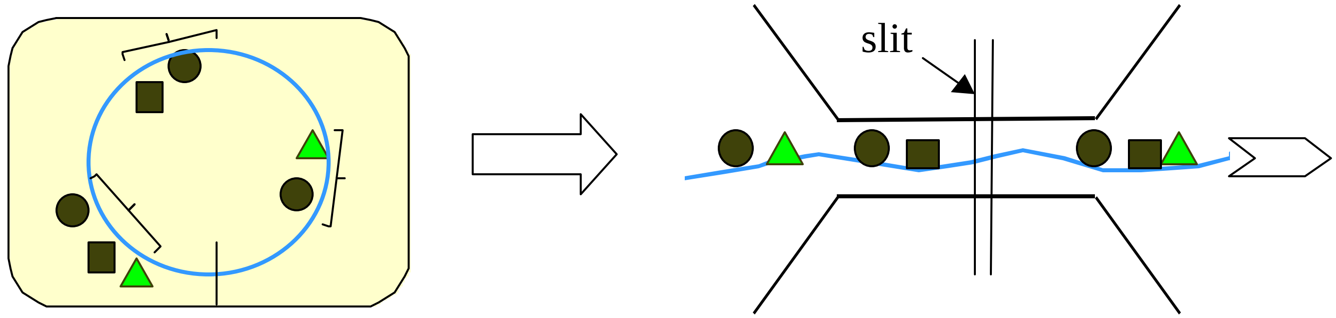
- Now want to find which TF bound to each gene

5. What transcription factors are on what genes?

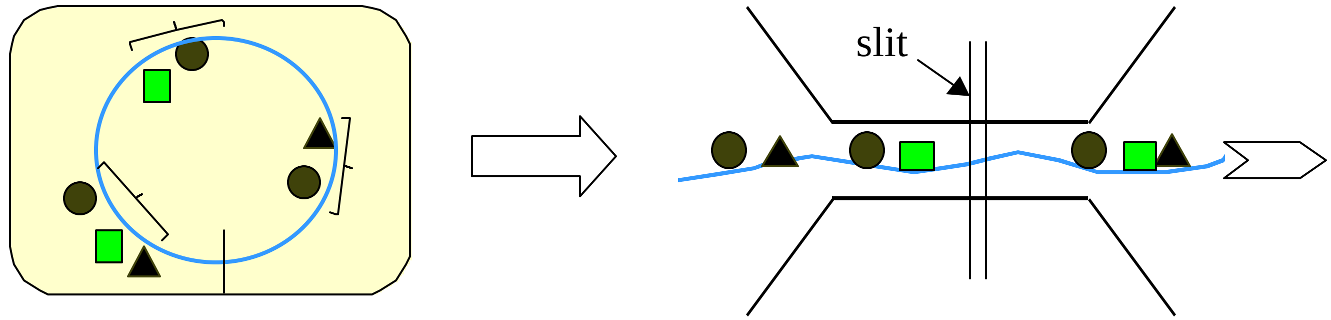
a. cleave DNA

b. near-field scanner

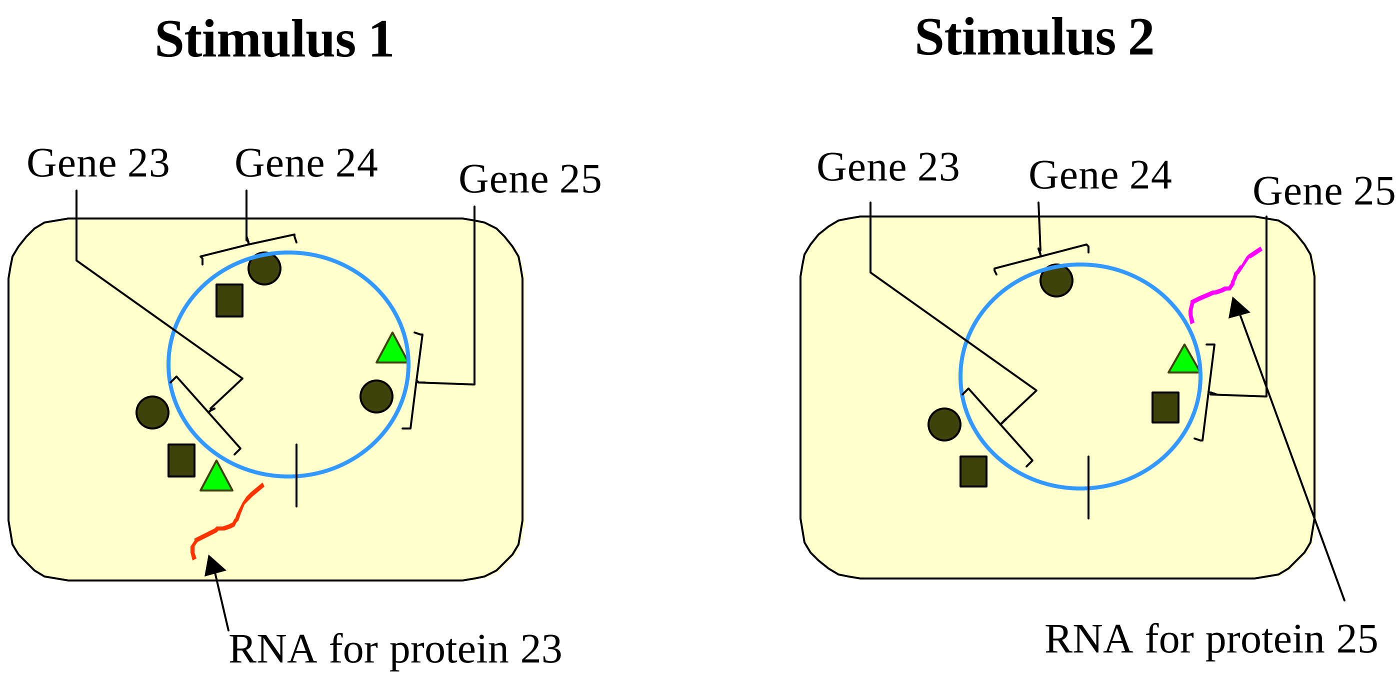
Strain 1 gives location of TF▲



Strain 2 gives location of TF■



6. Repeat for different stimuli => different genes expressed



- Now need to correlate genes produced with transcription profiles

7. Bioinformatics

		Transcription Factor				
		1	**2**	**3**	**4**	**…**
Gene Number	**1**	Δ	X	X	Δ	
	2	O	O	Δ	X	
	3	X	Δ	O	Δ	
	. . .	. . .	. . .	. . .	. . .	

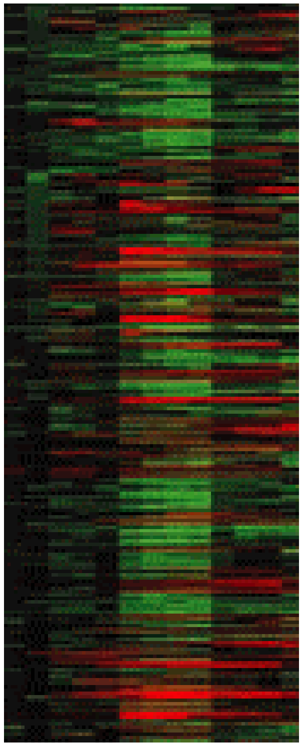
O: must have, D: irrelevant, X: must not have

- Use bioinformatic approach to discern required conditions of transcription for each gene: *What is the combinatorial code?*
- Seek higher level regulating algorithms: *Is there method to the madness?*

Gene Expression Analysis

Conditions

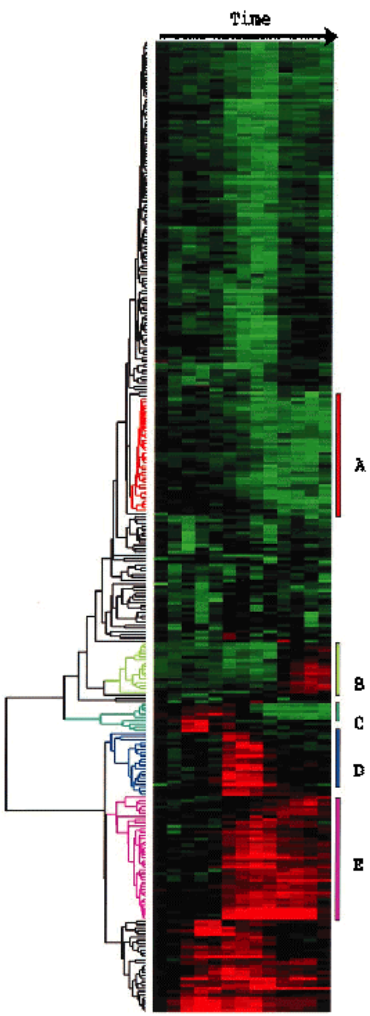
Genes



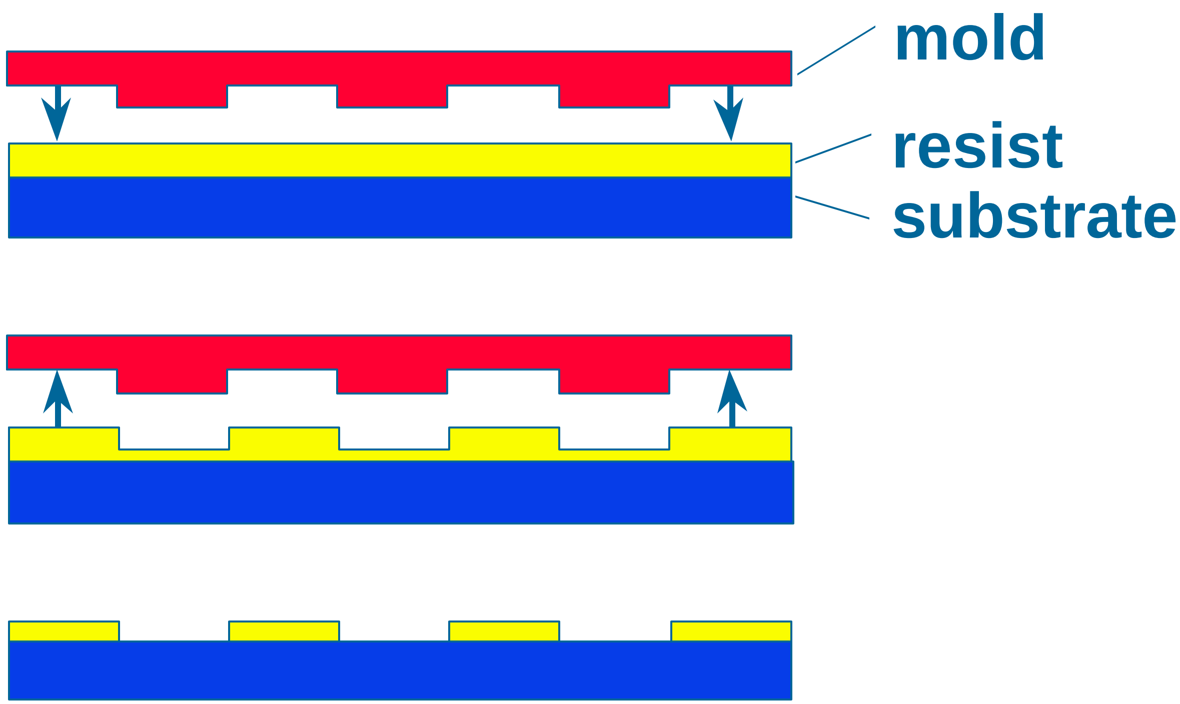
Gene Expression Analysis

Similar expression → **Similar function**

- **Genes are points in n-dimensional space**
- **Cluster genes**
  - **Hierarchial**
  - **k-means**
- **Linearly correlated genes**

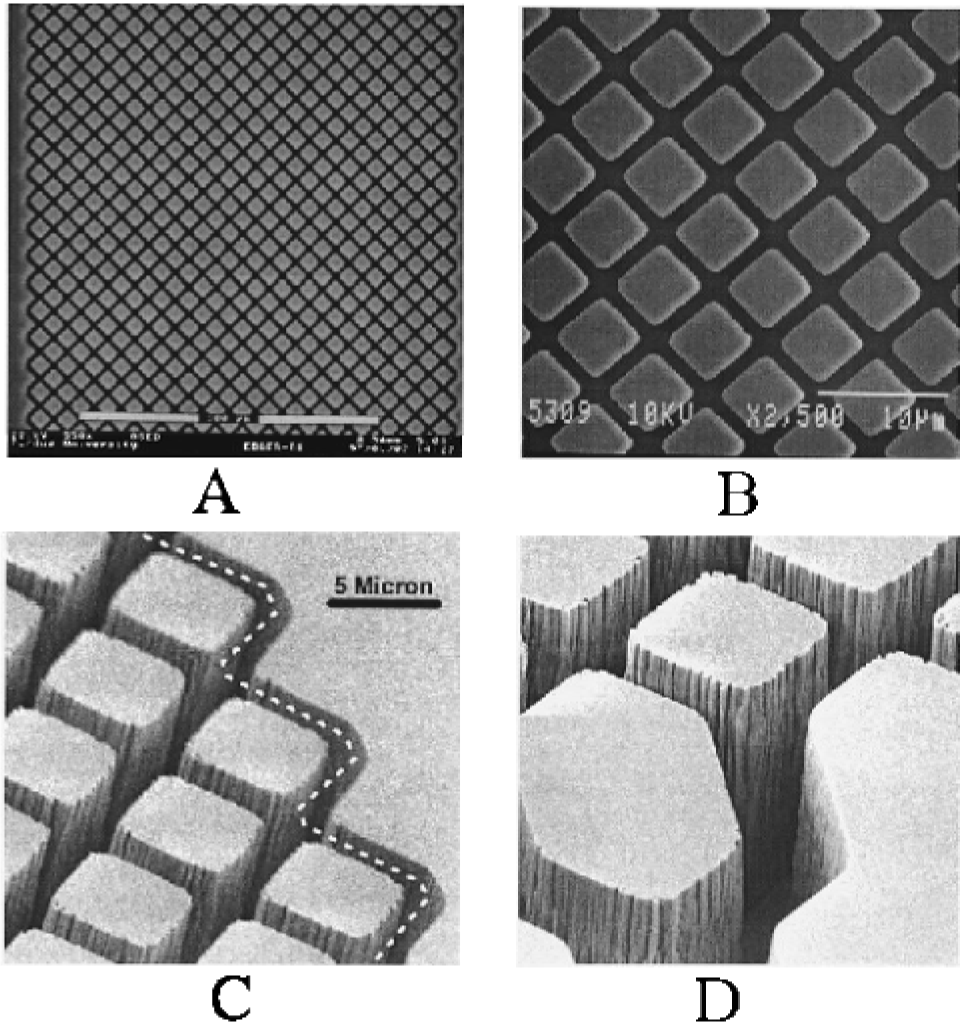


Patterning Technology: Nanoimprint Lithography



- Press Mold
- Remove Mold
- RIE

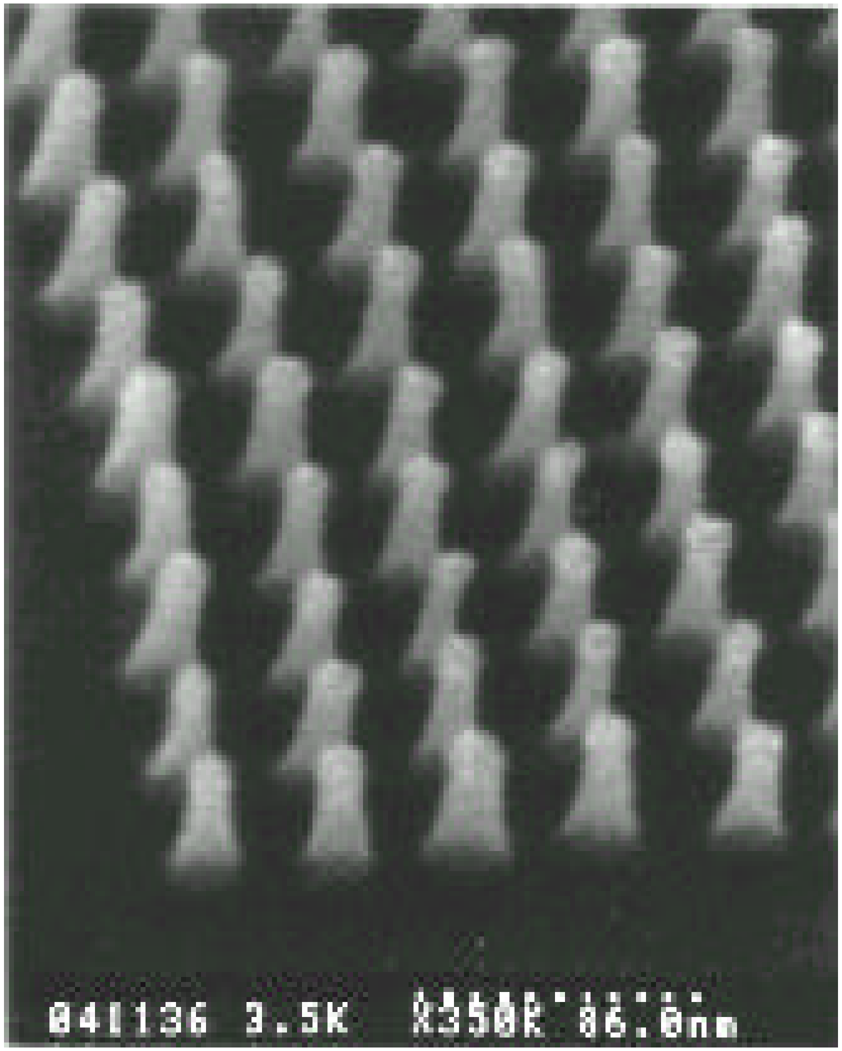
Chou, Krauss, and Renstrom, APL, Vol. 67, 3114 (1995); Science, Vol. 272, 85 (1996)



- **Small Structures over Large Areas**
- Electrophoresis and chromatography by artificial structures
- Smaller structures, differentiate smaller molecules …

  - Previous work by He, Tait and Regnier (*Anal. Chem.*, **70** (18), 3790 -3797, 1998)

Nano pillars



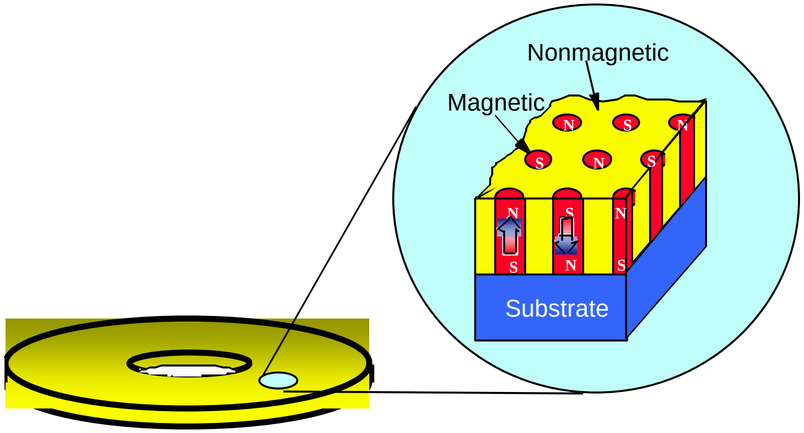
Height: 200 atoms

Width: 60 atoms

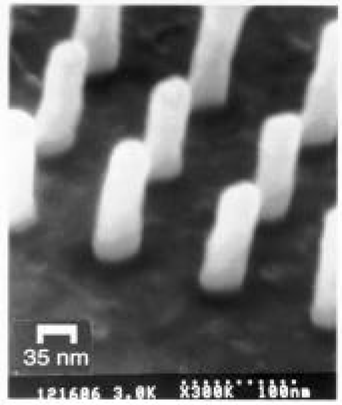
Nanopillar arrays

- 2-D sensor arrays
- Single domain magnetic particles
- Single electron particles

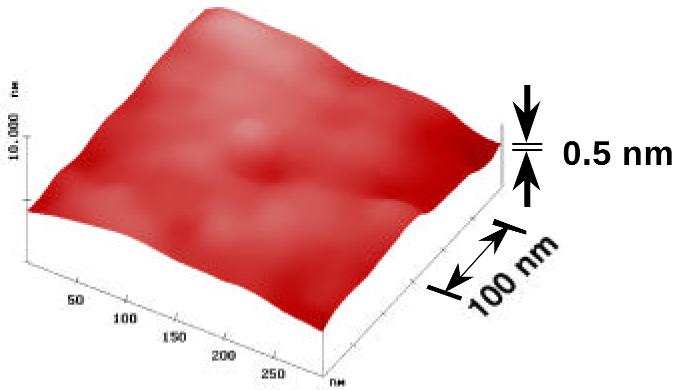
Nanomagnetic Structures for Bio-Applications



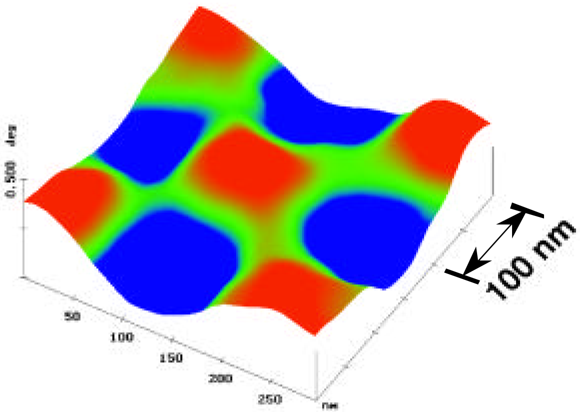
Schematic of planar quantum magnetic disk with perpendicular magnetization.



SEM micrograph of nickel pillars with nonmagnetic material removed.



AFM image of 3x3 bits of 65 Gbits/in^2 quantum magnetic disk.



MFM image of 3x3 bits of 65 Gbits/in^2 quantum magnetic disk.

NanoStructure Laboratory
PRINCETON UNIVERSITY

Some Big “If”s

- Single molecule resolution in near-field scanner
- Binding strength of transcription factors to DNA
- Scale-up issues (~5000 proteins in E-coli)
- There may be no combinatorial code
- Epigenetics
- ….

Two extreme models for transcription logic

- transcription as the ultimate Rube Goldberg; whatever works

- transcription as a well designed machine in which the complexity is programmed

Posters

Fluidic mechanisms in asymmetric arrays

Lotien Richard Huang, Robert Austin, James C. Sturm

Near-Field Scanning Device for DNA Analysis

Jonas Tegenfeldt, Robert H. Austin, Eugene Chan, Edward C. Cox

A Nanobrain in Bacterial Chemosensing

Mikhail N. Levit, Peter M. Wolanin , Jeffry B. Stock

Nanopatterning for the Analysis and Manipulation of Cellular and Sub-cellular Biological Systems

Han Cao, Paru Deshpande, Stephen Chou

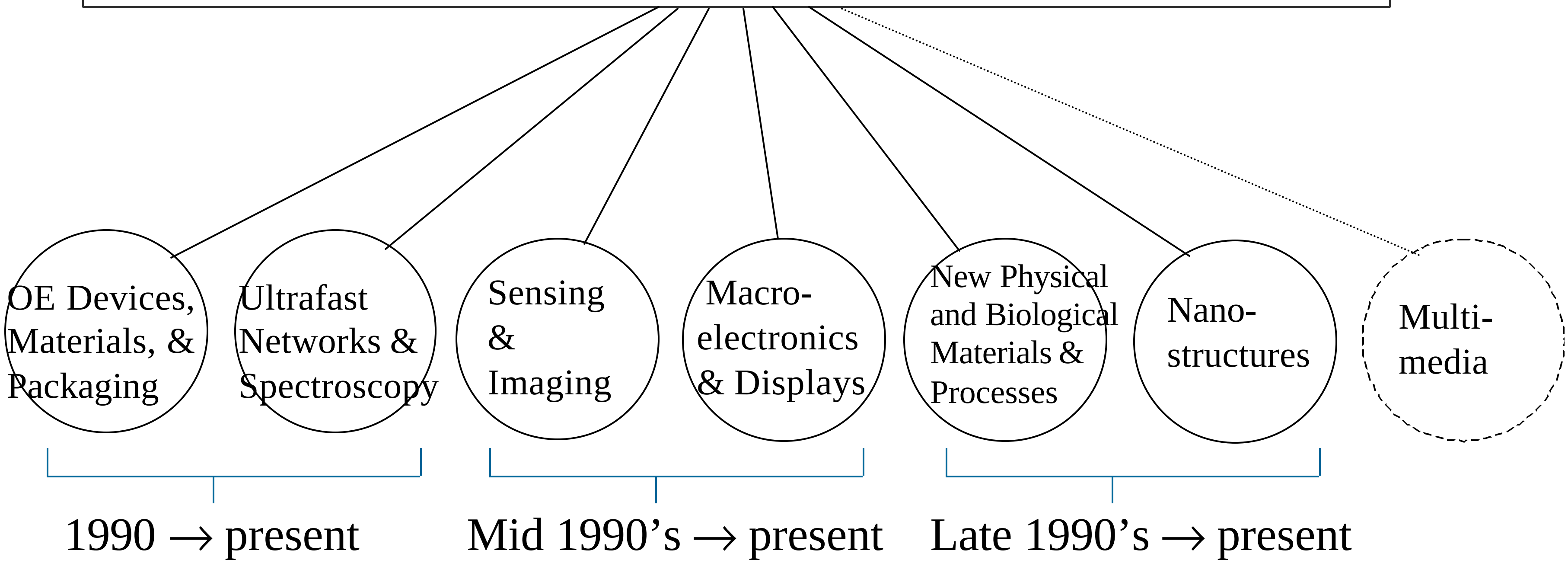
Deciphering the Transcriptional Code: Computational Questions and Approaches

Carl Kingsford, Mona Singh

Summary: Impact and Significance

- Understanding transcription is a central problem in understanding basic laws of how life works, both normal and diseased
- Basic laws of gene expression necessary to design future biological systems
- Combinatorial logic of gene expression may be a language of its own and shared by all organisms
- Multidisciplinary approach crucial to address problem

Princeton Center for Photonics and Optoelectronic Materials (POEM)



- Core strengths: photonics and nanostructures
- Mission: interdisciplinary, work with industry and government